DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health

Division of Research Resources
Biomedical Research Technology Program
Annual Progress Report
PART I, TITLE PAGE

1. PHS AWARD NUMBER: 5 U 4 1 R R 0 1 6 8 5 - 0 3

2. TITLE OF AWARD: BIONET NATIONAL COMPUTER RESOURCE FOR MOLECULAR BIOLOGY

3. NAME OF RECIPIENT INSTITUTION: IntelliGenetics, Inc.

4. HEALTH PROFESSIONAL SCHOOL (If applicable):

5. REPORTING PERIOD:

5a. FROM (Month, Day, Year): 0 3 - 0 1 - 8 6

5b. TO (Month, Day, Year): 0 2 - 2 8 - 8 7

6. PRINCIPAL INVESTIGATOR:

6a. NAME: Dr. Michael J. Kelly

6b. TITLE: President, IntelliGenetics

6c. SIGNATURE: Michael J. Kelly

7. DATE SIGNED (Month, Day, Year): 12-12-86

8. TELEPHONE (Include Area Code): 4 1 5 - 9 6 5 - 5 5 9 0

Part II. Description of Program Activities

This section of our Annual Report provides statistical information on the use of the BIONET^{tm} Resource. The period covered is 12/85 - 11/86, to coincide with the dates of preparation of our Report and to follow our procedure of providing a full year's statistical information to compare with previous years' Reports.

Individual sections are prepared under guidelines discussed previously with BRTP staff and used in our previous Reports. We use a format for reporting the hundreds of individual Principal Investigators' use that is easy for us to generate while retaining the critical information necessary for BRTP in its internal and governmental reporting requirements. Complete research abstracts are kept at IntelliGenetics and are available upon request.

The BIONET User community is divided into different classes, representing different levels of use of the computer system and staff resources, as follows:

- **Class I.** Class I users represent the Service component of the scientific community. They participate in the electronic communications facilities of BIONET (bulletin boards and electronic mail), and use the Core and Contributed Software libraries to pursue their research;

- **Class II.** Class II users represent the Collaborative component of the user community. Scientists in Class II enjoy all benefits of Class I use, and in addition contribute software and expertise to BIONET, working closely with BIONET staff. This category also includes bulletin board leaders, accounts by courtesy with other, related Resources (GenBank, NBRF/PIR, Dana Farber, etc.), National Advisory Committee members and accounts for communication with BIONET Satellites.

- **Class III.** The category of Class III access has been reserved for system managers of local computer facilities. Such persons might not qualify as Principal Investigators, but are willing to work closely with other researchers of BIONET at a local site to help them learn to use the system and telecommunications effectively. They share Class II privileges.

- **Class IV.** Class IV users consist of those scientists who wish access only to the electronic communication facilities of BIONET. They are given access to the electronic mail and bulletin board facilities.

Information on number of PI's by Class is summarized in Table II-1.

The total number of investigators with access to BIONET, 489, is about 40 less than the total presented in our last annual report. During the past year about 170 investigators chose not to renew their accounts primarily because of their anticipated lack of need for access to BIONET and the imposition of the subscription fee. During the year we have added about 130 new investigators to the Resource. We expect that the trend of bringing substantial numbers of new users on to the system will continue. Attrition in the future will not, however, be as heavy because all new users come to BIONET with full knowledge of

Table II-1: Summary of the BIONET User Community

Class I	444
Class II	37
Class III	4
Class IV	4
Total	489

the subscription fee. Rather, we expect that the growing use of locally-available sequence analysis software, provided through the BIONET Satellite program or PI's purchase of microcomputer software, will diminish the load per PI on BIONET, enabling us to support a larger number of users.

II.A. Scientific Subprojects

II.A.1. Collaborative Research and Service

In the following section we report the use of the BIONET Resource for Class I-IV users. The "Usage Factor" is reported as both central processor unit (cpu) time, in minutes and connect time in hours, for each Principal Investigator. These values are the sum of all usage by the PI and his or her group members ("Sub-I's"). We report data only on those PI groups that have used the Resource during the past 12 months. Of the 489 PI's, 418, representing about 1463 individual investigators, have accessed BIONET. Detailed statistics on the use by each individual are maintained by the BIONET computer and are available to interested parties.

We do not report Resource staff hours nor BRTP funds allocated for individual PI's because it is impossible to allocate these rationally to such a large user community. Summary information on allocation of staff hours is given in Section II.C, the Resource Summary Table.

II.A.2. Core Research and Development

We report on the standard form the summary information for our Core Research projects.

The Resource Technology used is the DEC-2060 computer for all projects.

The Usage Factor is reported as minutes of cpu time used for the project. This number is derived for each investigator by multiplying the fractional time ("FT") spent on the project times the total cpu time used during the past twelve months.

Resource Staff Hours are based on the same FT multiplied times the total hours spent on BIONET for the last twelve months.

"BRTP Funds Allocated" are calculated as followed, from the sum of the following components:

- Actual Personnel Costs. The personnel costs for each project are derived by multiplying the above FT for each BIONET staff person's time spent on the project times their respective annual salary plus fringe benefits; the actual personnel cost is the sum of these individual figures.

- Consultant Costs. The FT spent by a Co-Investigator involved in a project is multiplied times the total consulting cost for the Co-I; these are summed for each project where appropriate.

- Fraction of Awarded Funds. The fraction of total awarded funds for each project is derived by multiplying the fractional time spent on the project by the awarded funds (defined below). The fractional time is determined from the sum of hours spent on the project by all investigators divided by the sum of hours spent on BIONET by all investigators. In computation of awarded funds, we include the grant categories of *Supplies*, *Travel*, and *Other Expenses*. The categories of *Personnel* and *Consultants* are accounted for in the previous two computations. For this calculation we have used the actual time spent in the last twelve months and a cost basis of the estimated total expenditures in the above categories for this grant year. Although these are three months out of phase, we do not think the fractional time spent will change significantly during the next three months of the current grant period.

DRR SCIENTIFIC SUBPROJECT FORM

PART II, SECTION A — **INSTITUTION:** IntelliGenetics

AWARD NUMBER: 5 U 41 RR 01685-03 — **REPORT PERIOD:** March 1, 1986 to February 28, 1987

Fill out a separate Subproject Form for Core, Collaborative or Training. Check one of the following:

[X] CORE RESEARCH & DEVELOPMENT [ ] COLLABORATIVE RESEARCH & SERVICE [ ] TRAINING

Descriptive Title (80 characters) (1) / Abstract	(2) Science Code Axis I	(2) Science Code Axis II	(3) a. Investigator(s) Name (Last Name, First Name & Middle Initial) b. Department c. Non-Host Inst.	(4) USAGE FACTOR Resource Technology a/	(4) USAGE FACTOR CPU MIN USED	(4) USAGE FACTOR Resource Staff Hours	(5) BRTP Funds Allocated
Multiple Sequence Alignment Collection, Porting, and Evaluation of several computer programs for intercomparison of multiple (3 or more) biological sequences; Release of validated programs to the BIONET community with a review of strengths and weaknesses.	9	42, 68	a. Smith, Dennis H. Brutlag, Douglas L. Friedmann, Theresa A. b. IntelliGenetics	DEC-2060 " "	12 13 80	55 10 56	8526
BIONET Satellite Communications Districation of the computational and communication facilities of BIONET to other, geographically remote sites (Satellite Resources); linking these Satellites for rapid exchange of Electronic Mail and Bulletins, Data Files and Programs.	9	40, 42	a. Smith, Dennis H. Roode, David R. Relph, John R. Liebschutz, Robert Friedland, Peter Boyd, Brian Levy, Benjy b. IntelliGenetics	DEC-2060 " " " " " "	12 156 1289 137 3 1 1	55 180 257 196 10 27 39	48,434
Special Hardware for Text Search Exploring new parallel processing technologies for high speed text search and sequence homologies for applications to biological sequence retrieval and intercomparison.	9	42, 68, 70	a. Smith, Dennis H. Friedland, Peter Brutlag, Douglas L. b. IntelliGenetics	DEC-2060 " "	35 33 40	166 96 29	26,425
CUMULATIVE TOTALS: No. subprojects	3				1810	1176	83,385

a/ Identify Resource Technologies Used. b/ Give Hours Resource Technologies Used. See Instructions, page ii.

II.A.3. Training

We report summary information for our Training program. The sites at which BIONET provided some level of training are named here and are discussed in more detail in Chapter III, *Narrative Description*, section III.A.4

The method for calculation of Usage Factors and BRTP Funds Allocated is the same as that described above under Core Research and Development.

DRR SCIENTIFIC SUBPROJECT FORM

PART II, SECTION A **INSTITUTION:** IntelliGenetics

AWARD NUMBER 5 U 4 1 R R 0 1 6 8 5 — 0 3 **REPORT PERIOD:** March 1, 1986 **to** February 28, 1987

Fill out a separate Subproject Form for Core, Collaborative or Training. Check one of the following:

[ ] **CORE RESEARCH & DEVELOPMENT** [ ] **COLLABORATIVE RESEARCH & SERVICE** [X] **TRAINING**

Descriptive Title (80 characters) (1) **Abstract**	**(2) Science Code** Axis I	Axis II	**(3)** a. Investigator(s) Name (Last Name, First Name & Middle Initial) b. Department c. Non-Host Inst.	**(4) USAGE FACTOR** Resource Technology a/	CPU MIN USED	Resource Staff Hours	**(5) BRTP Funds Allocated**
BIONET Training Program Support of traing for BIONET scientists, including phone trainings, preparation of new documentation for training, and outside trainings at Stanford University, Miami Mid-Winter Symposia, American Scoiety of Biological Chemists, University of New Hampshire, and the Macromolecules, Genes and Computers meeting.	9	40.68	a. Allen, Marcia	DEC-2060	24	72	
			Bigham, Nancy	"	162	115	
			Brutlag, Douglas L.	"	66	48	
			Friedmann, Theresa A.	"	172	119	
			Friedland, Peter	"	3	10	
			Kristofferson, David	"	41	24	
			Lawler, Maryjo	"	164	221	
			Smith, Dennis H.	"	12	55	41,012
CUMULATIVE TOTALS: **No. subprbjects**	1				644	664	41,012

a/ Identify Resource Technologies Used. b/ Give Hours Resource Technologies Used. See Instructions, page 11.

II.B. Books, Papers, Abstracts

We report the publications by members of the BIONET scientific community on a version of the special form provided by BRTP. These publications have **ALL** arisen from use of BIONET, although support by BIONET and the NIH has not always been acknowledged.

The figures on *Cumulative Number Published* refer to the current year alone. This year we have received 114 publications resulting from the use of BIONET, a substantial increase over the 43 reports received last year.

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Cooke, N. E., David E., Vitamine D. Binding Protein is a Third Member of the Albumin and Alpha-Fetoprotein Gene Family. J. Clin. Invest. 76:2420-2424, 1985.

Cooke, N.E. Rat Vitamin D Binding Protein: Determination of the Full-Length Primary Structure from Cloned cDNA. J. Biol. Chem. 261:3441-3450, 1986.

Cooke, N. E., et al, Direct Regional Assignment of the Gene for Vitamin D. Binding Protein (Cc-Globulin) to Human Chromosome 4q11-4q13 and Identification of an Associated DNA Polymorphism. Human Genetics 73:335-229, 1986

Leibhaber, S.A., Ray, J., Cooke, N. E., Sythesis of Growth Hormore - Prolactin Chimeric Porteins and Processing Mutants by the Exchange and Deletion of Genomic Exons. J. Biol. Chem. 261:14, 301-14,306, 1986.

Elsholtz, Rosenfeld, et al, Two Different Cis-Active Elements Transfer the Transcription Effects of Both EGF and Phorbol Esters. Science In press.

Howe, C., Overton, C., et al, Expression of the Testis and Ovary is Regulated During Embryogenesis. In preparation.

Miller, R. H., Robinson, W., Common Evolutionary Origin of Hepatitis B Virus and Retroviruses. PNAS 83:2531-2535 1986

Keen, N., Tamaki, S., Structure of Two Pectate'yase Genes from Erwinia Chrysanthesis EC16 and Thier High Level Expression in Esckerichier Coli. J. Bacteriol. In press

Trumbly, R. J., Isolation of the Yeast CyC8 Gene Responsible for Catabolite Repression. In preparation.

Trumbly, R. J. Sequence Analysis of the Yeast CYC8 Gene Predicts a Highly Glutamine Rich Protein. In preparation.

Chen, L-S, Mullin, D., Newton A., Identification, Nucleotide Sequence and Control of Developmentally Regulated Promotors in the Hook Operon Region of Caulobacter crescentus. PNAS USA 83:3860-2864 1986.

Minnich, S., Newton, A., Promoter Mapping and Cell Cycle Regulation of Flagellin Gene Transcrpits in Caulobacter crescentus. PNAS USA In press.

Grimes, S. R. Jr., Daum H.A. III, Smith, J. M., et al A Histone H4 Gene is Closely Associated with the Testis-Specific H1t Gene. Nucl. Acids Resear. In preparation.

Strobel, M., Abelson, J. Intron Mutations Affect Splicing of S. Cervisiae SUP53 Precursor tRNA.
Molec. & Cell. Biol. 6:2674-2683 1986.

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Tschunko, A.H., Allen, S. L., Long Inverted and Direct Repeats Occur in the Vicinity of Eliminated Sequences of Tetrahymena thermophila. Nucl. Acids Resear. Submitted 1987.

Tschunko, A. H., Allen, S. L., Genomic Organization and Developmental Fate of Adjacent Repeated Sequences of Tetrahymena thermophila. Genetics Accepted 1987.

Lohe, A.R., Brutlag, D. L., Multiplicity of Satellite DNA Sequences in Drosophile Melanogaster. PNAS 83:696-700 1986.

Lohe, A. R., Brutlag, D. L., Adjacent Satellite DNA Segments in Drosophila : Structure of Functions. J. Mol. Biol. Accepted 1987.

Lohe, A. R., Brutlag, D. L., Identical Satellite DNA Sequences in Sibling Species of Drosophila. J. Mol. Biol. In press 1987

Pritchard, A.E., Seilhamer, J., Cummings, D., Paramecium mitochondrial DNA Sequences and RNA Transcripts for Cytochrome Oxidase Subunit I, RF1, and Three ORFs Adjacent to the Replication Origin. Gene 44:243-253 1986.

Mahelingam, R., Cummings, D., et al Identification of Paramecium mitochondrialproteins using antibodies Raised Against Fused Mitochondrial Gene Products. Gene In press 1986.

Davis, R.J., Johnson, D.J., Czech, M.P., et. al. Identification of Serine 24 as the Unique Site on the Transferrin Receptor Phosphorylated by Protein Kinase C. J. Biol. Chem. 261 9034-9041 1986

Krawetz, S.A., Connor, Wayne, Cannon, Paul D., and Gordon H. Dixon A Vector-Primer-Cloner-Sequencer Plasmid for the Construction of cDNA Libraries: Evidence for a Rat Glyceraldehyde-3-Phosphate Dehydrogenase-Like mRNA and a Ferritin mRNA within Testis. DNA 5:427-435.

Lovett, M., cheng, Z.Y., Epstein, C.J. et al Molecular Markers for the Agouti Coat Color Locus of the Mouse. Genetics In press.

Lincoln, J.E. Cordes, S., Read, E., and Fischer, R.L. Regulation of Gene Expression by Ethylene during Fruit Development. PNAS In press

Glimcher, L.H. and Griffith, I.J. Mutation of Class II MHC molecules. Immunology Today In press

Griffith, I.J., Choi, E., and Glimcher, L.H. A single mutation in a Class II A ∂ gene alters T cell recognition. Proc. Nat'l Acad. Sci. In press

Griffith, I.J., Carland, F.M., and Glimcher, L.H. Alteration of a non-polymorphic residue in a Class II Eß gene eliminates an

COLLABORATIVE RESEARCH AND SERVICE

antibody-defined epitope without affecting T cell recognition. Submitted.

Biggs, J., Searles, L.L., and Greenleaf, A.L. Structure of the Eukaryotic Transcription Apparatus: Features of the Gene for the Largest Subunit of Drosophila RNA Polymerase II. Cell 42:611-621. 1985

Hamori, E. Novel DNA Sequence Representations. Nature 314: 585-586 1985

Hamori, Eugene and Zhou, Bing Analysis and Comparison of Two Fully Sequenced Viral DNA Genomes Using H Curves. Biophysical Journal 49, 125a 1986

Hamori, Eugene Analysis of Long DNA Sequences Using the Novel Method of H Curves (Abstr.) Hungarians in Science p. 768 Budapest, 1986 (MTA/MTESZ)

Hamori, Eugene Long Range Nucleotide-compostion Patterns in Fully-sequenced Genomes. Biol. Chem. Hoppe-Seyler 367S, 226 1986

Herrick, G., Cartinhour, S., Dawson, D. et al Mobile Elements bounded by C4A4 telomeric repeats in Oxytricha fallax. Cell 43:759-768 1985

Herrick, G., Cartinhour, S., Williams, K.R., and Kotter, K.P. Multiple Sequence Versions of the Oxytricha fallax 81-MAC alternate processing family. J. Protozool. In press.

Bray, S.J. & Hirsh, J. The D.virilis dopa decarboxylase gene is developmentally regulated when integrated in D. melanogaster. EMBO J. 5:2305-2312 1986

Scholnick, S., Bray, S.J., Hirsh, J., et al Central nervous system and hypoderm regulatory elements of the D. melanogaster dopa decarboxylase gene. Science In press

Morgan, D., Johnson, W.A. & Hirsh, J. Regulated splicing produces different forms of dopa decarboxylase in the central nervous system and hypoderm of Drosophila melanogaster. EMBO J. In press

Wiginton, D.A., Kaplan, D.J., Hutton, J.J. et al Complete Sequence and Structure of the Gene for Human Adenosine Deaminase. Biochemistry In press

Wabiko, H., Raymond, K.C., Bulla, L.A. Bacillus thuringiensis Entomocidal Protoxin Gene Sequence and Gene Product Analysis. DAN 5:305-314 1986

Aird, S.D., Kaiser, I.I., Lewis, R.V., Kruegel, W. Complete Amino Acid Sequence for the Basic Subunit of Crotoxin. Arch. Bich. and Biophys 249:296-30 1986

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Aird, S.D., Kaiser, I.I., Lewis, R.V. and Kruegel, W. Rattlesnake Presynaptic Neurotoxins: Primary Structure and Evolutionary Origin of the Acidic Subunit. Biochem. 24:7054-7058 1985

Chung-Her Jenh, Tiliang Deng, Johnson, Lee F., et al Mouse Thymidylate Synthase. Locations of Introns and Multiple Transciptional Start Sites. J. Biol. Chem In press 1986

Donohue, T. J., Hoger, J., and Kaplan, S., Cloning and Expression of the Rhodobacter sphaeroides Reaction Center H. Gene. J. Bact. 168: In press 1986

Donohue, T. J., McEwan, A., Kaplan, S. Cloning, DNA Sequence and Expression of the Rhodobacter sphaeroides Cytochrome c2 Gene. J. Bact. 168: In press 1986

Kemper, B., Molecular Biology of Parathyroid Hormone. Critical Rev. in Biochem. 19:353-380 1986

Govind, S., Bell, P.A., Kemper B., Structure of Genes in the cytochrome P-450PBc subfamily: Conservation of Intron Locations in the Phenobarbital-inducible Family. DNA 5: In press 1986

Hussain, M., Pastor, F., Lampen J.O., Cloning and Sequencing of blaz Encoding B-LactamaseIII, a Lipoprotein of Bacillus cereus 569/H. J. Bact. In press 1987

Boyer, P.D., Lengyel, J.A., Molecular Characterization of bsg25D: A Blastoderm-specific Locus in Drosophila melanogaster encoding a Product with Domains of Structural Similarity to Fos and Tropomyosin Proteins. Submitted

Wernke, S.M., Lingrel, J. B., Nucleotide Sequence of the Goat Embryonic ∂ Globin Gene (zeta) and Linkage and Evolutionary Analysis of the Complete ∂ Globin Cluster. J. Mol. Biol. 192: In press 1986

Hind K.R., Litman G. W., Major Reorganization of Immunoglobulin VH Segmental Elements During Vertebrate Evolution. Nature 320:546-549 1986

Kimmel, B. E., Yarbrough, L. R., et al, Tubulin Genes of the African Trypanosome Trypanosoma brucei rhodesiense: Nucleotide Sequence of a 3.7 kb Fragment Containing Genes for Alpha and Beta Tubulin. Gene 35:237-248 1985

Hirschberg, R., Yarbrough., L. R., et al, Cloning and Characterization of the Nitrogenase Genes from Anabaena veriabilis. J. Biotech. 2:23-38 1985

Wu, J., Yarbrough, L. R., et al, Tubulin Genes of the African Trypanosome: Nucleotide Sequence and Conservation of 5' Untranslated Regions of Tandem Repeats. Submitted

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Marsh, J., Eveleth, Evidence for Evolutionary Duplication of Genes in the Dopa Decarboxylase Region of Drosophila. Genetics 114:469-483 1986

Bland, M.M., Levings, C. S. III, and Matzinger, D. F., The Tobacco Mitochondrial ATPase Subunit 9 Gene is Closely Linked to an Open Reading Frame for a Ribosomal Protein. Mol. Gen. Genet. 204:8-16 1986

Miller, V. L., Mekalanos, J. J., Genetic Analysis of the Cholera Toxin-Positive Regulatory Gene toxR. J. Bact. 163:580-585 1985

Betley, M, Miller, V., Mekalanos, J. J., Genetics of Bacterial Exterotoxins. Ann. Rev. Microbio. 40:577-605 1986

Osawa, S., Miller, T. B. Jr., et al, Isolation of Partial cDNA's for Rat Liver and Muscle Glycogen Phosphorylase Isozymes. FEBS 202:282-288 1986

Oshima, R. G., Millan, J., Cecna, G., Comparison of Mouse and Human Keratin 18: A Component of Intermediate Filaments Expressed Prior to Implantation. Differentiation. In press

Bikenmeyer, L., Sugisaki, H., Ray, D. S., The Majority of Minicircle DNA in Crithidia fasciculata Strain CF-C1 is of a Single Class with Nearly Homogeneous DNA Sequences. Nucl. Acids Res. 13:7107-7118 1985

Birkenmeyer, L., Ray, D. S., Replication of kDNA in Kinetoplasts Isolated from Crithidia fasciculata: Identification of Minicircle DNA Replication Intermediates. J. Biol. Chem. 261:2362-2368 1986

Krusius, T., Ruoslahti, E., Primary Structure of an Extracellular Matrix Proteoglycan Core Protein Deduced From Cloned cDNA. PNAS USA 83:7683-7687 1986

Argraves, W., Pytela, R., Ruoslahti, E., et al, cDNA Sequences from the Alpha Subunit of the Fibronectin Receptor Predict a Transmembrane Domain and a Short Cytoplasmic Peptide. J. Biol. Chem. 261:12922-12924 1986

Suzuki, S., Ruoslahti, E., et al, cDNA and Amino Acid Sequences of the Cell Adhesive Protein Receptor Recognizing Vitronectin Reveal a Transmembrane Domain and Homologies with Other Adhesion Protein Receptors. PNAS USA In press

Bourdon, M., Shiga, M., Ruoslahti, E., Identification from cDNA of the Precursor Form of a Chondroitin Sulfate Proteoglycan Core Protein. J. Biol. Chem. 261:12534-12537 1986

Bourdon, M. , Shiga, M., Ruoslahti, E. Gene Expression of the Chondroitin Sulfate Proteoglycan Core Protein PG19. Mole. Cell. Biol. In press

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Poncz, M., Schwartz E., et al, Cloning and Characterization of Platelet Factor 4 cDNA Derived from a Human Erythroleukemic Cell Line. Blood 69: In press 1987

Sykes, P., Sokatch, J. R., et al, Cloning and Expression of Branched Chain Keto Acid Degydrogenase in Pseudomonas Putida and Escherichia Coli. Submitted 1986

Bova, C., Manfredi, J., Swanstrom R., env Genes of Avian Retroviruses: Nucleotide Sequence and Molecular Recombinants Define Host Range Determinants. Virology 152:"343-354 1986

Lerner, C., Switzer, R. L., Cloning and Structure of the Bacillus Subtilis Aspartate Transcarbamylase Gene (pyrB). J Biol. Chem. 261:11156-11165 1986

Theil, E. C., Ferritin: Structure, Gene Regulation and Cellular Function in Animals, Plants and Microorganisms. Ann. Rev. Biochem. 56: In press 1987

Roth, W., Wahba, A., et al, The Expression of a Gene for Mouse Eukaryotic Elongation Factor Tu During Murine Erythroleukemic Cell Differentiation. In preparation

Faber, S., Wilcox K., Association of the Herpes Simplex Virus Regulatory Protein IÇ∏4 with Specific Nucleotide Sequences in DNA. Nuc. Acids Res. 14:6067-6083 1986

Kuroda, M., Yanofsky, C., et al, Regulatory Elements Common tothe Bacillus Pumilus and Bacillus Subtilis trp Operons. J. Bact. 167:792-798 1986

Stanton, L., Schwab, M., Bishop, J.M., Nucleotide Sequence of Human N-myc. PNAS USA 83:1772-1776 1986

Gerondakis, S., Bishop, J.M., Structure of the Protein Encoded by the Chicken Proto-Oncogene c-myb. Mol. Cell. Biol. In press 1986

Ness, D. B., Grumet F.C., New Polymorphisms of HLA-B27 and Other B Locus Antigens Detected by RFLP Using a Locus Specific Probe. Human Immun. In press 1986

Wells, D., Bains, W., Kedes, L., Condon Usage in Histone Gene Families of Higher Eukaryotes Reflects Functional Rather than Phylogenetic Relationships. J. Mol. Evol. 23:224-241 1986

Erba, H., Gunning, P., Kedes, L., Nucleotide Sequences of the Human Upsilon Cytoskeletal Actin mRNA: Anomalous Evolution of Vertebrate Non-muscle Actin Genes. Nucl. Acid Res. 14:4275-5294 1986

Wade, R., Kedes, L., et al, Nucleotide Sequence, Tissue Specific Expression, and Chromosome Location of Human Carbonic Anhydrase

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

III: The Human CAIII Gene is Located on the Same Chromosome as the Closely Linked CAI and CAII Genes. PNAS USA In press 1986

Regier, J. C., Evolution and Higher-order Structure of Architectural Proteins in Silkmoth Chorion. EMBO J. 5:1981-1989 1986

Duffy, L., Kurosky, A., Lai, C-Y., Cholera Toxin A Subunit: Functional Sites Correlated with Regions of Secondar Structure. Arch. Bio. Biophys. 235:549-555 1985

Nagel, G., Kurosky, A., et al, Evidence for the Expression of Three Genes Encoding Homologous Atrial Gland Peptides that Cause Egg Laying in Aplysia. J. Biol. Chem. 261:7853-7859 1986

Singh, S., Kurosky, A., Awasthi, Y., Human Liver Glutathoine S-transferase: Chemical Characterization and Secondary Structure Comparison with Other mamalian Glutathione S-transferases. Biochem. J. Submitted 1986

Yamanaka, M., Tobin, S., et al, Structure and Expression of the Drosophila Calmodulin Gene. Submitted 1986

Vigoreaux, J., Tobin, S., The 5C actin Gene of Drosophila melanogaster Encodes Multiple Transcrpits Sharing the Entire Protein Coding Sequence but Differing in their 5' and 3' Non-translated Sequences. In preparation 1987

Flink, I., Morkin, E., et al, Complete Amino Acid Sequence of Human Thyroxine Binding Globulin Deduced from Cloned DNA: Close Homology to the Serine Antiproteases. PNAS USA 83:7708-7712 1986

Bishop, D., Calhoun, E., et al, Human a-galactosidase A: Nucleotide Sequence of a cDNA Clone Encoding the Mature Enzyme. PNAS USA 83:4859-4863 1986

Lindahl, L., Zengel, J., Ribosonel Genes in Escherichia coli. Ann. Rev. Genet. 20:297-326 1986

Edenberg, H., Zhang, K., Isolation of a Mouse-Liver Alcohol Dehydrogenase cDNA. ICSU Short Reports 1:162-163 1984

Edenberg, H., Zhang, K., et al, Cloning and Sequencing of cDNA Encoding the Complete Mouse-Liver Alcohol Dehyrogenase. PNAS USA 82:2262-2266 1985

Crabb, D., Edenberg, H., Cloning and Sequencing of a cDNA Encoding Rat Liver Alcohol Dehydrogenase. Alcohol and Alcoholism In press 1986

Crabb, D., Edenberg, H., Complete Amino Acid Sequence of Rat Liver Alcohol Dehydrogenase Deduced from the cDNA Sequence. Gene In press 1986

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Leung, W., Liu, A., et al, Protein Engineering of Human Interferon Gamma by Using Novel Approaches for Chemical-Enzymatic Gene Synthesis and Fragment Replacement Method for Sit Specific Mutagenesis. Biol. of Interferon Sys. p. 412-424 1985

Liao, S., Leugn, W., Kanamaru, T., Structure and Epitope Specificity of Human Melanoma-Associated Oncofetal Antigen gP87 Analyzed with Monoclonal Antibodies. J. Dermat. 12:117-128 1985

Hasnain, S., Manavathu, E., Leung, W., DNA-mediated Transformation of Chlamydomonas Reinhardi Cells: Use of Amino-glycoside 3'-phosphotransferase as a Selectable Marker. Mol. and Cell. Biol. 5:3647-3650 1985

Hasnain, S., Leugn, W., et al, Development of Host Cell System for Production of Proteins of Biological and Medical Importance: Implications for Biotechnology. "Proteins in Food, Health and Industry" In press 1986

Jing, G., Liu, A., Leung, W., A Method for the Preparation of Size Marker for Synthetic Oligonucleotides. Analy. Biochem. 155:367-378 1986

Diaz-Mitoma, F., Leung, W., et al, Detection of Epstein Barr Virus in Throat Washings by DNA-DNA Dot Hybridization. J. Infec. Dis. In press 1986

Laksman, E., Hiruki, C., Wu, X., Leung, W., Use of [32p]-RNA Probes for the Dot-hybridization Detection of Potato Spindle Tubor Viroid. J. Virol. Meth. In press 1986

Hasnain, S., Manavathu, E., Leung, W., Construction of a Plasmid Vector Bearing the Initiation ATG Condon for Direct Expression in E. coli. Plasmid In press 1986

Leung, W., Manavathu, E., et al, Development of Fungal and Algal Cells for Expression of Foreign Genes. Applied Virology 2: Academic Press 1986

Kolbe, H., Wohlrab, H., Sequence of the N-terminal Formic Acid Fragment and Location of the N-Ethylmaleimide-binding Site of the Phosphate Transport Protein from Beef Heart Mitochondria. J. Biol. Chem. 260:15899-15906 1985

Rasmussen, V., Wohlrab, H., Bovine Cardiac Mitochondrial ADP/ATP-Carrier: Two Distinct mRNAs and an Unusually Short 3'-Noncoding Sequence. Biochem. Biophys. Res. Comm. 138:850-857 1986

Upton, C., McFadden G., DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species. Mol. Cell. Biol. 6:265-276 1986

Upton, C., McFadden G., Tumorigenic Poxviruses: Analysis of Viral DNA Sequences Implicated in the Tumorigenicity of Shope Fibroma Virus and Malignant Rabbit Virus. Virology 152:308-321 1986

Part II Section B Award Number: 5U41RR01685-03
INSTITUTION: IntelliGenetics Report Period: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH AND SERVICE

Upton, C., McFadden, G., Identification and Nucleotide Sequence of the Thymideine Kinase Gene of Shope Fibroma Virus. J. Virol. In press 1986

Chang, W., McFadden, G., et al, Genome of Shope Fibroma Virus, a Tumorigenic Poxvirus, Contains a Growth Factor Gene with Sequence Similarity to those Encoding Epidermal Growth Factor and Transforming Growth Factor Alpha. Mol. Cell. Biol. In press

Upton, C., Carrell, R., McFadden, G., A Novel Member of the Serpin Superfamily is Encoded on a Circular Plasmid-like DNA Species Isolated from Rabbit Cells. FEBS Lett. In press

CUMMULATIVE # SUBMITTED		BOOKS 0	PAPERS	17	ABSTR.	0
CUMMULATIVE # PUBLISHED		BOOKS 0	PAPERS	67	ABSTR.	3
CUMMULATIVE # IN PRESS		BOOKS 0	PAPERS	26	ABSTR.	1

II.C. Resource Summary Table

The Resource Summary Table includes the totals from the previous sections of *Core Research and Development* and *Training*. The totals for *Collaborative Research and Service* we derived as the sum of the following components:

- **Usage Factor**. These are computed as the differences between the totals for the BIONET minus the Core R+D and Training figures. The total for *CPU Min. Used* is the sum of all CPU time consumed by BIONET users and staff. The total of staff hours is self explanatory.

- **BRTP Funds Allocated**. This is computed as the difference between the total budget for BIONET minus the categories of Core R+D and Training, and minus the capital equipment expenditures for the year ($8000).

The category of *Collaborative Research and Service* includes an entry of $153,100 in the column *Other Funds*. This is the total money collected over the period 12/85 - 11/86 for subscription fees. Each PI is asked to pay an access fee to help defray the telecommunication costs for access to BIONET; this fee is currently $400/year. By agreement with BRTP, these access fees are not grant related income.

The balance of these fees carried forward from the previous year (as of 12/1/85) was $25,000. After twelve additional months of collecting subscription fees and disbursing them for telecommunication expenses, the balance is now $120,100. BIONET has increased the number of telecommunication ports by four to provide improved service. This action has increased the telecomunication charges by a factor of two which should result in a total expenditure for telecommunications of approximately $200,000 during the next twelve months. At this rate we have judiciously reserved the current balance to serve our needs over the next calender year and plan on substantially depleting this current balance by November of 1987 in accordance with the timing of the receipt of the subscription fees from individual users.

The category of *Administration/Miscellaneous* includes only the Usage Factor of BIONET's share of the DEC-2060 system overhead accounts. No facility staff computer time, work hours, or BRTP funds are allocated; we consider such time and funds to be an integral part of the support of the other components of the Resource. We do include as Funds Allocated the $8000 to purchase items of capital equipment.

The category of down time includes the sum of scheduled and unscheduled maintenance on the DEC-2060 computer. In the period 12/85 - 11/86, there was a total of 99 hours (5935 cpu minutes) of downtime:

- 3266 cpu minutes of scheduled downtime for preventive maintenance and several system-related tasks including major time for installation of Version 6.1 of the TOPS-20 operating system.

- 2669 minutes of downtime were due to unscheduled maintenance.

The downtime reported in the Summary Table is 50% of the total, reflecting BIONET's access to 50% of

the machine. Note that the (total) unscheduled maintenance of 2669 minutes is only 0.5% of the total cpu time available. Considering both categories of downtime, the machine has been available for use by BIONET scientists, 99.5% of the time, 24 hours a day, seven days a week. No funds have been allocated to this category.

PART II, SECTION C RESOURCE SUMMARY TABLE

AWARD NUMBER: 5U41RR01685-03

REPORT PERIOD: March 1, 1986 to February 28, 1987

RESOURCE COMPONENT	(1) Number Subproject	(2) Number Publications	(3) Number Investigators	(4) USAGE FACTOR Resource a/ Technology	CPU Min Used	Resource Staff Hrs	(5) BRTP Funds Allocated $	(6) Resource Fees $ Collected	(7) Other Funds $
CORE RESEARCH & DEVELOPMENT	3	-0-	Staff as Above	DEC-2060	1810	1176	83,385		
COLLABORATIVE RESEARCH & SERVICE	489	114	1709	DEC-2060	7990	9283	538,235		153,100
TRAINING	N/A	N/A	Staff as Above	DEC-2060	664	664	41,012		
ADMINISTRATION/ MISCELLANEOUS	SEE PREFACE TO THIS		SECTION	DEC-2060	8466		8,000		
DOWN TIME					2968				
GRAND TOTALS	492	114	1709		24,845	11,123	670,632		

a/ Identify Resource Technologies Used. b/ Give Hours Resource Technologies Used. See Instructions, page iv.

Part III. Narrative Description

III.A. Summary of Research Progress

We have had a very successful year of operation. Previous problems in funding were resolved, and our operation has stabilized to the great benefit of the scientific community. Although our funds were cut 8% at the beginning of the year, we were able to anticipate those cuts and our budget was supplemented in October 1986 by funds transferred to the NIH from the National Science Foundation to help support telecommunications. This support, plus the continuing support from our Principal Investigators in the form of subscription fees for access to BIONET, has allowed us to focus on the scientific goals of the Resource rather than on fiscal dislocations. As a result, we have been able to provide high quality, uninterrupted service to the BIONET community and devote additional personnel resources to Collaborative and Core Research projects. A possible cloud on the horizon is a lack of resolution of the indirect cost rate. This could seriously affect our future performance.

The following sections describe in detail our accomplishments in the several components of the BIONET Resource. Here, in brief, are some of the most notable:

- The community in terms of active PI's has grown substantially in the past year. To serve better this community, we have increased the number of communication ports into BIONET. The statistics on the community were described in the introduction to Chapter II. The new communication facilities are described under *Resource Facilities*, below;

- The community published or has in press more than 114 papers during the past twelve months, an indicator of the crucial role BIONET plays in computational support of research in molecular biology;

- The response of the central DEC-2060 computer during prime time has become quite slow, due to the high level of use by the community. We are addressing this problem through installation of additional BIONET Satellites and through mechanisms for distributing time consuming computations to other machines (see *Core Research* below).

- We have taken several steps to improve the electronic communications available through BIONET, including substantial work on revamping the bulletin board system (see *Collaborative Research* below), and installation of mail forwarding among facilities via Telenet and ARPANET (see *Core Research* and *Resource Facilities* below);

- We have begun the collection and dissemination of several programs designed to help solve the very important problem of alignment of multiple biological sequences. This important Core Research activity will place BIONET in a leadership role in making such programs routinely available (see *Core Research* below).

- We have identified several vendors for special text searching hardware and identified two as providing the most promising machines. Such machines have the potential for revolutionizing methods for text search and calculation of sequence homologies. We are currently making arrangements for access to these machines (see *Core Research* below).

III.A.1. Service

The Service component of BIONET includes primarily Class I investigators who use the BIONET Core and Contributed program Libraries (see section III.A.5.c) to support their research. They have access to all functions of the System Library (section III.A.5.d) as well, but use primarily the systems for electronic communication (electronic mail and bulletin boards).

The Service component also includes Class III investigators who are given access based on their responsibilities for computing in molecular biology using department, school or campus-wide computer facilities. As part of their agreement for membership in BIONET, they provide information about the resource and access to it for their local community. There are currently four investigators in this category:

- Robert Gross - Dr. Gross has responsibilities in the newly-created Molecular Genetics Center at Dartmouth. He facilitates local use of BIONET, and has helped arrange a training at Dartmouth (see *BIONET Training Program* below);

- Kenneth Manly - Dr. Manly is a Cancer Research Specialist at the Roswell Park Memorial Institute at SUNY-Buffalo. He provides local support for BIONET users at this institution;

- Pavel Vitek - Dr. Vitek is Head of Computer Facilities at the Imperial Cancer Research Fund in London. He provides local support and serves as a source of information on BIONET for other UK investigators;

- Charles Lawrence - Dr. Lawrence is Principal Investigator for the Molecular Biology Information Resource, a regional resource funded by BRTP/DRR/NIH. He provides information about BIONET to the regional community supported by his Resource.

The final category of Service users is Class IV, a Class provided for those investigators who wish access only to the communication facilities of BIONET. There are currently four investigators in Class IV: Marlene Belfort, New York State Dept. of Health at Albany; Dieter Soll, Yale University; Jean Walat, BIOSIS; and Edward Hoover, Colorado State University.

III.A.1.a. Scientific Consulting: Class I, III and IV Support

The Service component of the BIONET Resource is supported by a group of BIONET Scientific Consultants. The Consultants interact with the community in a variety of ways, including direct support via telephone calls, electronic mail and terminal links with individual investigators. Support is also provided through their participation at major meetings, trade shows, and trainings. The consultants also provide on-line and printed documentation for User Manuals, program descriptions and system procedures.

We currently have two full-time and one half-time Consultants. They provide direct support to the community 50% of their time. The other 50% is devoted to participation in Core and Collaborative Research projects as described in subsequent sections.

III.A.1.b. Service

The Service component of the BIONET Resource includes primarily Class I investigators and takes the form of answering questions by phone, by electronic mail, and by terminal links. A survey of the monthly phone, mail and terminal links for the past year shows the different uses of the BIONET Resource.

The monthly inquiry rates for the five categories of Programs and Databases, TOPS20 System, PCs and PC Software, Telecommunications, and BIONET Administration are listed below.

Table III-1: Summary of Monthly Rates of Inquiries

Category	Number of Inquiries	Percent of Total Inquiries
Programs and Databases	153	36
TOPS20 System	117	27
Telecommunications	72	17
Administration	45	10
PC and PC Software	41	10
	----	----
TOTALS	428	100

As can be seen from Table III-1, the largest number of inquiries concern the use of BIONET's programs. This first category, Programs and Databases, has been subdivided into five different scientific and program categories in Table III-2.

As shown in Table III-1, the largest number of questions received by the BIONET Staff concern the use of the resource's programs. The analysis in Table III-2 shows that the majority of these inquiries relate to the use of the databases and the database access programs. The sequence entry/manipulation and sequence analysis programs categories were a fairly distant second at 18% and 15% respectively, and the rest of the program categories accounted for less than 10% each of the total rate of inquiries. Multiple-Sequence alignment inquiries mostly concerned the use of William Bains' contributed XMULTAN program. TOPS20 system program inquiries mainly concerned the use of FIND and XSEARCH, and experiment planning and analysis inquiries covered the use of the SIZER, MAP and CLONER programs. The category OTHER includes inquiries on the use of other contributed programs, such as Michael Zuker's BIOFLD, and other miscellaneous questions.

Returning to Table III-2, 27% of user questions concerned the TOPS20 operating system, specifically the manipulation of files and directories, the use of the text editors, the control of output to the terminal, and the access of programs.

Table III-2: Summary of Monthly Rates of Questions for Programs and Databases.

Category	Number of Inquiries	Percent of Total Inquiries
Database Searches and Databases	65	42
Sequence and Gel Data Entry and Manipulation	27	18
DNA and Protein Sequence Analysis	23	15
Multi-Sequence Alignment	11	7
TOPS20 System Programs	10	7
Experiment Planning and Analysis	5	3
Other	12	8
	----	----
TOTALS	153	100

The third largest category in Table III-1 is Telecommunications. This category includes both inquiries concerning the procedures involved in connecting to the BIONET computer and the quality of the communications between remote users and BIONET.

The last two categories in Table III-1, BIONET Administration and PCs and PC software, each accounted for 10% of the inquiries. The BIONET administration category consisted mostly of application requests, training session information, and manual requests. The majority of the BIONET administration

calls were routed directly to the BIONET Administrator and are not included in Table III-1. The questions most frequently asked about PC software pertained to file transfer and terminal emulators. Questions of this nature mainly concerned software which runs on PCs manufactured by IBM and Apple.

The majority of the users' questions were answered immediately by our Scientific Consultants. Even the more difficult questions usually received a response within a day. The availability of the Scientific Consultant staff resulted in substantial savings of investigator research time.

III.A.1.c. Scientific Case Studies Using BIONET

After examining the large number of publications received, we chose the following two examples of research which utilized the BIONET resource.

"Evolution and High-Order Structure of Architectural Proteins in Silkmoth Chorion" EMBO Journal 5:1981-1989, 1986, J.C. Regier.

Jerome C. Regier joined BIONET in February 1985 and has used over 1400 CPU minutes and 158 connect hours during the past year on BIONET.

Dr. Regier has been studying the structure of silkmoth chorion with the aim of understanding eukaryotic morphogenesis at the molecular level. Silkmoth chorion morphogenesis involves the temporally regulated production of more than 100 follicle cell-specific proteins in widely varying amounts. The chorion proteins assemble to form two predominant types of structures: lamellae that are highly ordered helicoidal arrays of fibrils and filler that is a sponge-like network. Filler accounts for roughly 5% of the chorion's mass and consists of only two proteins E1 and E2. It forms hollow breathing channels through lamellar chorion and molds a small number of outer surface lamellae into crown-shaped structures called aeropyle crowns. Prior to this report, the E1 cDNA and genomic clones had been sequenced. Their predicted secondary structures and hydropathicity profiles revealed a periodicity postulated to have functional significance. No homology was detected between E1 and lamellar sequences.

Dr. Regier's group sequenced genomic and cDNA clones that encode the E2 silkmoth chorion protein. E2 was found to have two distinct domains with the amino terminal domain consisting of four alternating stretches of hydrophobic and hydrophilic residues, the first three of which are homologous in sequence to about half of the E1 protein. The carboxyl terminal domain of E2 is much longer. It is hydrophilic and consists entirely of multiple tandem copies of a single, variant hexapeptide repeat sequence that is absent from E1.

Dr. Regier searched for homology between the nonrepetitive region of E2, which spans residues 1-127, and the sequences found in the three large databases that are available on BIONET: the National Institutes of

Health DNA sequence library (GenBank), the European Moleculary Biology Laboratory DNA sequence library, and the National Biomedial Research Foundation's protein sequence database. This search was performed using the IFIND program. E1 gave the highest similarity score of any sequence in all three databases and was the only sequence whose score did not form part of the continuum of other scores.

He also searched for similar lysine- and asparagine-like hexapeptide repeats in other sequences, using BIONET's QUEST program. For the protein searches, he looked for at least three tandem repeats of "Lys Lys Asp ..." or four of "Asn .....", where a "." represents any amino acid. None were found. For the nucleotide searches, he looked for coding regions that contained at least three tandem repeats of "AA.AA.GA. ... ..." or four of "AA[T or C] ... ..." For the first octadecanucleotide, no sequences were found. For the second, four groups of closely related sequences were found that could be translated in the appropriate open reading frame. One (plasmodium surface antigen) encoded multiple tetrapeptide repeats of Asn-Ala-Asn-Pro. The second (trypanosome ribosomal protein) contained long stretches of polyasparagine. The last two (drosphila GAG-like protein and yeast positive regulatory protein) contained no evidence of being within a repeat region other than for the presence of the four asparagine codons. He concludes that the lysine and asparagine repeats of E2 have not been found in other sequences.

Based on the above sequence information, he suggests that the most likely evolution for the E genes is that the ancestral E gene encoded an amino terminal sequence similar to that in the modern-day E1 and E2. After the ancestral gene duplicated *in toto*, one of the copies (the E2 gene ancestor) added on the hexapeptide repeat domain. This could have occurred either by reduplicating a sequence already present within the gene or by transposition of an unrelated sequence. He theorizes that the absence of an intron at the border of the repeat domain argues against the exon shuffling hypothesis. They are currently sequencing E genes from another silkmoth species in hopes of better understanding the origin of the repeat domain.

"Intron Mutations Affect Splicing of *Saccharomyces Cerevisiae* SUP53 Precursor tRNA", Molecular and Cellular Biology, 6: 2674-2683, 1986, M.C. Strobel and J. Abelson.

John Abelson has been a BIONET user since December 1984 and has used over 600 CPU minutes and 206 connect hours of BIONET computer time in the past year. The Abelson lab also actively uses BIONET's electronic mail and bulletin board facilities to further their research efforts.

Drs. Strobel and Abelson investigated the role of intron structure and sequence on precursor tRNA splicing and subsequent mature tRNA production. In their experiments, they used mutants of the *saccharomyces cervisiae* amber suppressor tRNA gene SUP53 that encodes a pre-tRNA containing a 32 base intervening sequence. They constructed two types of mutants: a first type with an internal deletion

of the natural SUP53 intron and a second type with a novel intron. These mutant genes were transcribed *in vitro*, and the end-processed transcripts were analyzed for their ability to serve as substrates for the partially purified *S. cerevisiae* tRNA endonuclease and ligase. After integration of these mutant genes into the yeast genome, the *in vivo* suppressor tRNA function of these mutant alleles was correlated with the *in vitro* phenotype.

The *S. cerevisiae* pre-tRNA introns have a common structure where each intron folds to form an extension of the anticodon stem. The splice junctions are generally in single-stranded loops and are a constant distance apart (6 base pairs). The extended anticodon stem is stabilized by base-pairing between the anticodon and the intron, allowing the stem nucleotides to be highly base-stacked while the loop nucleotides are not.

The secondary structures of SUP53 and the mutant RNAs were determined by using the RNA sequence folding program of Michael Zuker which is available on BIONET. These structures were determined both as part of the total pre-tRNA and as subfragments spanning the anticodon stem-loop region (including intervening sequences).

Drs. Strobel and Abelson's experiments showed that certain intron-mutated precursor tRNAs were refractory to the tRNA-splicing endonuclease both *in vitro* and *in vivo*. Furthermore, secondary mutations at the 3' intron-exon junction could partially rescue the nonspliced phenotype *in vitro* and *in vivo*.

Single-nucleotide changes were shown to have a drastic effect on the predicted secondary structure of the intron. The results were consistent with the idea that correct, efficient splicing is dependent on the intron's ability to form a stable extension of the anticodon stem and retain the 3' splice junction in a single-stranded region. The BIOFLD program on BIONET was used to predict the secondary structure of each precursor. Precursors whose secondary structures formed a stable extended stem were readily spliced. In contrast, two unspliced precursors exhibited severe distortion of the extended stem. Additionally, while tRNA precursors exhibited no sequence conservation at exon-intron junctions or within introns, certain primary sequences at these junctions were thought to be inappropriate.

III.A.1.d. Comments from the Community

A measure of the success of BIONET is the response from the community about the Resource. As part of our reapplication procedures for renewal of access to BIONET, we include a section for comments and suggestions. These comments and suggestions are a valuable source of information to us in order to improve BIONET. In this section we summarize the comments and suggestions received during the current cycle of renewals (timed to coincide with preparation of our Annual Report). See Appendix I for a copy of the reapplication form.

It is clear from the comments received that BIONET is a valuable part of all research efforts entailing nucleic acid or protein sequence analysis. The most valued aspect of the computer resource is the ability to search the major databases. Fifty five percent of those responding cited this reason. The most commonly used search is for sequence homologies. Other capabilities cited included the usefulness of the programs for the assembly and analysis of sequence data. Although the majority of the work on BIONET involves nucleic acid sequence analysis, it is interesting that about 25% of the citations specifically mentioned the use of BIONET for protein sequence analysis. In addition, it was gratifying to note that 15% of the repondents cited the usefulness of the bulletin boards in their research efforts. If the citations concerning the electronic mail system are included, approximately one quarter of the respondents mentioned the use of the BIONET communications facilities. Although this usage is still below that of the sequence analysis programs on BIONET, one should keep in mind that bulletin boards and electronic mail are still rather novel concepts in the biological sciences. When viewed in this light, the figures show that the acceptance of electronic communications by the research community is off to a good start.

The comments also detail the need for further improvement of the BIONET resource in several areas. We have already begun to address these remaining points of user disatisfaction even before receiving the reapplication comments.

The main problem cited in the comments section remains the slow response time and heavy usage of the time-shared DEC 2060 system. We are taking several steps to address this problem. In addition to our efforts to reduce the user load through the BIONET Satellite Program (section III.A.3.b below), we took immediate action to encourage use of the batch job facility on the 2060. Batch jobs allow users to run CPU-intensive tasks during off-peak hours. Our efforts included (1) a thorough rewriting and simplification of the documentation describing the batch technique; (2) dissemination of information about the technique via numerous bulletin board notices and login banner messages; and (3) individual assistance offered to users by our consulting staff. In most instances the consultants wrote the first batch job control file for investigators who then proceeded to use it as a model for further efforts on their own.

The second most often cited complaint concerned the uneven quality of the program documentation. With the addition of new staff at BIONET we have already been able to address part of this problem through the addition of new on-line help menus (see section III.A.1.f below and Appendix II). We have also begun work on a new manual of examples. This should enable users to find rapidly the appropriate sequence of commands required to accomplish their research tasks. These examples will first be presented on-line to obtain user feedback and will then be made available in printed format after they meet with user approval. The first of these examples will go on-line during the month of January 1987.

The recent switch to the TELENET telecommunications network caused some degree of confusion and

dissatisfaction, but we have taken action to address individual phone connection problems as they have arisen. As a result the number of problems of this nature is diminishing (see section III.A.5.a)

Several investigators expressed concern about the extended period of time between the publication date of new sequence information and its availability in the GenBank and EMBL databases. We are developing new procedures to allow BIONET users to contribute their sequences for on-line use as soon as the data is ready for submission to the major databanks. This project is described in section III.A.2.d below and should commence during January 1987.

The remainder of the comments involved individual suggestions for additional features on the resource and we are considering each suggestion carefully. Several of these requests have already been implemented. These recent changes include the addition of the Brookhaven protein databank and better editors on the system. Requests for more graphical program output, and dot matrix homology search programs are in the process of being addressed.

Finally, we note that only two investigators out of the 171 reapplicants complained about the cost of the subscription fee indicating that this is no longer a major area of concern.

III.A.1.e. PC/BIONET Communications - Distribution of the Resource

It has been clear from the beginning of the operation of BIONET that the majority of the user community had access to personal computers, and that they were looking for ways to use the PC's effectively in conjunction with BIONET. For example, sequences may be entered using a PC and stored on a floppy disk without the need to connect to BIONET. The file can be transmitted to BIONET later over the telephone lines. This reduces the load on the system. We have strongly supported this method of access, to the extent of maintaining a lending (and on-line) library of software and documentation for file transfer and terminal emulation programs. The growing availability of PC-based software for sequence analysis is another means by which the burden on the DEC-2060 will be reduced significantly.

To further facilitate the use of PCs, the IBM-PC public domain version of BIONET's on-line EMACS editor has recently been promoted on the system. "MicroEMACS" is distributed free to users and has the virtues of producing ASCII text files compatible with the BIONET software and of utilizing essentially the same command set as the mainframe editor. This should facilitate sequence entry on user PC's. The consultants have also provided assistance in converting files into a format compatible with the on-line software. They are actively encouraging long-time BIONET users to switch from the older SOS line editor to EMACS. The availability of an on-line interactive tutorial for learning EMACS, as well as recent improvements in on-line documentation for both EMACS and PC-to-mainframe file transfers, should promote this change.

BIONET has also sought to standardize file transfer protocols on the system by vigorously promoting the use of the public-domain Kermit software. Several tests conducted by the BIONET staff during the transition from UNINET to TELENET proved once again that Kermit was the most trouble-free of the file transfer protocols. An installation program has been added to the IBM-PC version of Kermit in the BIONET lending library. This assists scientists in their initial use of the software.

III.A.1.f. System Help Utilities

To facilitate use of BIONET, information about the system has been organized in a series of menus which users can view by typing "HELP ME" at the system level. A copy of the main menu accessed by this command is provided in Appendix II. The topics listed on this menu include (1) a guide to the IntelliGenetics programs and contributed software on the system; (2) instructions for using the electronic mail and bulletin board systems; (3) how to locate other users on the system; (4) operating system help; (5) on-line database information; (6) batch job instructions; (7) how to obtain consulting help; (8) how to use the on-line text editors; (9) how to transfer files between microcomputers and mainframes; (10) and assistance for TELENET. An upcoming feature on the main menu will be an index which will refer to new on-line program usage examples.

Several of the menu topics are broken down further into subtopics in lower-level menus (see Appendix II). This organization allows the user to remember only a single command, "HELP ME", to find his/her way around the system. This command is reinforced continually by a login banner reminder. After the program examples are implemented, we anticipate that any remaining difficulties in learning the use of the BIONET resource will be completely resolved.

In the few months that the HELP ME menu has been available, it has rapidly become the most frequently used help file on the system, being accessed at an average rate of just over 300 times a month. The consulting staff has received numerous favorable comments about the menu system during telephone conversations with users and also in electronic mail messages. With the coming addition of on-line program examples, HELP ME should become even more useful to the BIONET community.

III.A.2. Collaborative Research

BIONET's collaborative community is made up of several components, encompassing efforts by outside scientists working in conjunction with BIONET staff. In subsequent paragraphs we discuss each component in more detail:

- **DEC-2060 Software Contributors.** This component includes those persons who have contributed software for use by the BIONET community on the central DEC-2060 computer;

- **PC-Based Software.** This component includes our efforts to gather and disseminate PC-based software of special utility to the community.

- **Data Contributors**. This component includes those persons who, together with BIONET staff, contribute data useful to the community;

- **Liaison with Other Resources**. Several accounts have been established to promote sharing of information among molecular biology computing resources;

- **Bulletin Boards**. This component involves scientists who have agreed to maintain bulletin boards of special interest to the BIONET community.

III.A.2.a. DEC-2060 Software Contributors

The primary efforts of the collaborative research this year have gone into helping make programs developed by BIONET users available to others. This has included (1) conversion of a number of programs from other languages and operating systems to the BIONET environment; (2) altering the programs to accept BIONET file formats; (3) setting up bulletin boards announcing programs available for downloading to microcomputers; and (4) preparing lending libraries of the programs which can be physically sent out to BIONET users.

The following sections describe each of the programs that we have made available on BIONET, either as a program on the BIONET computer or as a program that can be executed on a local microcomputer. For programs that are used on BIONET we can provide a direct measure of their use. For those programs used on microcomputers we can only provide the number of times that the program file has been referenced (downloaded) and the number of independent requests for a copy of the program via the lending library.

DFASTP and DFASTN - BIONET and IntelliGenetics have made substantial efforts to bring up the DFASTP and DFASTN programs from Dr. Bill Pearson and to make them available to the entire community. IntelliGenetics ported the code for Pearson's programs from his VAX version to TOPS-20 versions (XFASTP and XFASTN) at no cost to BIONET. These programs are used heavily to search the PIR protein or the NIH GenBank databanks for homology with newly determined sequences. XFASTN and XFASTP are substantially faster than the IFIND program (which is based on the original Wilbur and Lipman NUCALN algorithm) and help reduce the total amount of computer time that BIONET uses in such searches.

BIONET users have performed 373 searches of the current PIR protein database using XFASTP. This amounts to 97 searches per month or over 3 searches per day. Unfortunately this is far less than the use of IFIND by BIONET users (346 searches per month or over 10 searches per day) searching the same database. The current BIONET documentation describes how to use the IFIND program but the use of the contributed XFASTP is only documented on line and in bulletin board messages. We can stimulate use of XFASTP by further messages on BBOARDS. Since IntelliGenetics has licensed the XFASTP and XFASTN programs from Dr. Pearson for incorporation into their IFIND program, the difference between

CPU usage of IFIND and XFASTP will be eliminated upon the release of the new versions of the core software.

XMULTAN - XMULTAN is a program developed by Dr. Bill Bains for aligning multiple homologous DNA sequences. While it can only align sequences which are at least 60% homologous, it is an extremely rapid program and expands the capabilities of the resource. Multiple sequence alignment is useful for BIONET users studying evolution and for those trying to obtain a consensus from many sequences of similar function. Appendix III shows seven related satellite DNA sequences isolated from several sibling species of *Drosophila* which were aligned with XMULTAN in less than 1 minute of CPU time on BIONET.

XMULTAN was modified by BIONET to be compatible with our file system and a menu driven front end was added to make the program easier to use. The original version was a non-interactive batch oriented program. A number of BIONET users have successfully used XMULTAN and provided us with valuable feedback. Although available for only two months, it is currently being used at a rate of 65 times a month.

RNAFOLD and BIOFLD - Last year Dr. Michael Zuker, from the NRC laboratory in Ottawa Ontario made BIOFLD available as a program on the BIONET computer. This program predicts RNA secondary structures and has been used 509 times for an average rate of 24 times per month.

ALIGN - Dr. Dan Davison, while a graduate student at Stony Brook at SUNY wrote and contributed two versions of his ALIGN program. The first version runs directly on the BIONET computer. It can be used to align two very long DNA sequences including those which are not very homologous to each other and contain large gaps. The alignments are significantly better than similar heuristic alignments obtained from the SEQ SEARCH procedure in the BIONET core programs. The alignments are not as good as obtained from the SEQ ALIGN procedure, but Dr. Davison's ALIGN program is significantly faster. The ALIGN program has been used 155 times (27 times per month average).

IDEAS - While at the NIH Dr. Kanehisa contributed this package of software to BIONET and made the program compatible with ours. The usage of this package has decreased this year to about once per month for two reasons. First, much of the functionality of these programs is already contained within the IntelliGenetics core library, and secondly, Dr. Kanehisa is now at Kyoto University and has interacted less with the resource due to communications costs.

XPROF - Dr. George Rose at Pennsylvania State University has contributed the DEC-VAX Fortran version of his method for calculating hydropathicity profiles for proteins based on empirical observations on the extent to which amino acid residues are found to be exposed or buried. We have modified this program to run on the DEC-2060.

A number of new collaborations are currently underway to bring even more functionality to the BIONET system. This includes obtaining a number of programs requested by BIONET users as well as a number of new relevant databases.

New Multiple Sequence Alignment Programs - XMULTAN is primarily useful for alignment of nucleic acid sequences since homologous proteins are generally less than 60% identical. A number of programs have been described in the literature for performing multiple alignments of protein sequences. Dr. Joel Sussman of the Weizmann Institute has made available his VAX Fortran program for aligning three protein sequences simultaneously. This program is a straightforward extension of the Needleman-Wunsch method for aligning two protein sequences and is currently being converted for use on BIONET. Unfortunately, it will be limited to aligning only three sequences at a time, and like the original Needleman-Wunsch procedure, it is a CPU and memory intensive application.

IntelliGenetics is also developing a new program for multiple sequence alignment of proteins and DNA sequences in collaboration with Dr. Hugo Martinez of the University of California, San Francisco. This program (GENALIGN) will be available in January of 1987. GENALIGN is a very rapid program for multiple alignment of either protein or DNA sequences and should adequately fill the need for both multiple protein sequence alignment and for consensus DNA alignments at low levels of homology.

Phylogenies And Evolutionary Trees - A suite of programs for constructing phylogenies based on sequence relationships has been written by Dr. Joe Felsenstein, Department of Genetics, University of Washington. Dr. Felsenstein has made these programs, originally written for a VAX computer, available to us. BIONET is currently modifying them for use on our resource. This program, coupled with the multiple sequence alignment programs mentioned above will provide the essential tools to carry out evolutionary biology studies with BIONET.

Finally, we expect that connection to the ARPANET will make a large number of libraries of public domain software readily available to BIONET users. One of the IntelliGenetics staff already maintains a collection of Macintosh public domain software obtained largely from the SUMEX computer resource at Stanford University. With the pending connection of the BIONET computer to the ARPANET the following software collections will be also available: <INFO-IBMPC> at ISI; <INFO-KERMIT> at COLUMBIA; <INFO-MAC> at SUMEX; <PC-BLUE>, <CPM>, <UNIX> and <MS-DOS> all at SIMTEL-20, <IBM> at DEC Marlboro etc. We will support access to these compilations of programs by publishing bulletins and procedures for accessing these computers via file transfer protocols. We will also maintain lists of programs available at each of these sites. With the advent of these resources directly available on ARPANET we will probably stop local maintenance of the <KERMIT> and <MACINTOSH> directories.